

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2000, 21:41:12 ; Search time 1895.15 Seconds  
(without alignments)  
-951.699 Million cell updates/sec

Title: US-08-988-242-1\_COPY1232\_1825  
Perfect score: 594  
Sequence: 1 CAGGTACAGCTAACGGCTT.....AAACGAGCTCGTGCGGATA 594

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 821193 seqs, -1518192014 residues  
Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0  
Maximum DB seq length: 1000000  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

- GenEmbl:\*
- 1: gb\_bal:\*
- 2: gb\_ba2:\*
- 3: gb\_on:\*
- 4: gb\_ov:\*
- 5: gb\_pat:\*
- 6: gb\_ph:\*
- 7: gb\_pl1:\*
- 8: gb\_pl2:\*
- 9: gb\_pr1:\*
- 10: gb\_pr2:\*
- 11: gb\_pr3:\*
- 12: gb\_ro:\*
- 13: gb\_sts:\*
- 14: gb\_sy:\*
- 15: gb\_un:\*
- 16: gb\_vt:\*
- 17: em\_fun:\*
- 18: em\_hum1:\*
- 19: em\_hum2:\*
- 20: em\_in:\*
- 21: em\_on:\*
- 22: em\_or:\*
- 23: em\_ov:\*
- 24: em\_pat:\*
- 25: em\_ph:\*
- 26: em\_pl:\*
- 27: em\_ro:\*
- 28: em\_sts:\*
- 29: em\_sy:\*
- 30: em\_un:\*
- 31: em\_vt:\*
- 32: gb\_htg1:\*
- 33: gb\_htg2:\*
- 34: gb\_in1:\*
- 35: gb\_in2:\*
- 36: gb\_ba1:\*
- 37: gb\_ba2:\*
- 38: em\_hum3:\*
- 39: em\_hum4:\*
- 40: gb\_pr4:\*
- 41: gb\_htg3:\*
- 42: gb\_htg4:\*
- 43: gb\_htg5:\*
- 44: gb\_htg6:\*

594  
-1232  
-----  
1825

- 45: gb\_htg7:\*
- 46: em\_htg1:\*
- 47: em\_htg2:\*
- 48: em\_htg3:\*
- 49: em\_hum5:\*
- 50: gb\_pl3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	594	100.0	3402	5	AR047920	Sequence 1
2	590.8	99.5	3402	5	A48910	U24190 Trypanosoma
3	590.8	99.5	3402	35	TCU24190	AL022022 Mycobacte
4	45	7.6	47852	1	MTV023	L27277 Micrococcu
5	44	7.4	2402	1	MLURHO	AJ132828 Spermatoz
6	43.2	7.3	2028	7	SS1132828	M30933 E.tenella a
7	42.8	7.2	767	34	EIMMAX	D14486 Equine herp
8	41.6	7.0	5880	16	HSE4	U87960 Rattus norv
9	41.2	6.9	2479	12	RNU87960	AF020275 Homo sapi
10	40.4	6.8	477	11	AF020275	AL133030 Homo sapi
11	40.4	6.8	3932	10	HSR801297	X83546 R.norvegicu
12	40.4	6.8	3941	12	RNLAR2	AL009198 Mycobacte
13	40.4	6.8	69350	1	MTV004	AP000556 Homo sapi
14	40.4	6.8	149618	10	AP000556	AP000557 Homo sapi
15	40.4	6.8	150036	10	AP000557	AP000552 Homo sapi
16	40.4	6.8	157086	10	AP000552	AC007957 Homo sapi
17	40.4	6.8	158529	44	AC007957	AC009516 Homo sapi
18	40.4	6.8	169237	40	AF139019	AF139019 Cepaea ne
19	40	6.7	624	35	CRGCR1	X17207 Chlamydomon
20	39.6	6.7	618	7	CRGCR1	AF087653 Homo sapi
21	39.4	6.6	1442	40	AF087653	L38713 Gallus gall
22	39.2	6.6	818	4	CHKPROTAMI	M28100 Gallus gall
23	39.2	6.6	1528	4	CHKPROTAMN	X95503 M.musculus
24	39.2	6.6	2771	12	MMZINCFPR	X95504 M.musculus
25	39.2	6.6	3732	12	MMZINCFIP	AB018491 Homo sapi
26	39	6.6	259	9	AB018491	AB007820 Homo sapi
27	38.8	6.5	720	9	AB007820	X12735 Barley Cab-
28	38.8	6.5	1030	7	HVCAB2	X68361 M.fascicula
29	38.8	6.5	2858	10	MFAPOA4A	M15764 D.melanogas
30	38.4	6.5	150	34	DROPNFS1	M62760 Homo sapien
31	38.4	6.5	2268	9	HUMCOUPII	X61945 D.melanogas
32	38.4	6.5	2722	34	DMNG2	AB010711 Mus sp. m
33	38.4	6.5	2837	12	AB010711	AF030376 Acidamino
34	38.4	6.5	3015	2	AF030576	M23222 D.melanogas
35	38.4	6.5	5632	34	DROFSHB	M23221 D.melanogas
36	38.4	6.5	7792	34	DROFSHA	AC013031 Drosophil
37	38.4	6.5	11196	43	AC013031	AC014003 Drosophil
38	38.4	6.5	37518	43	AC014003	AL024484 Drosophil
39	38.4	6.5	39565	34	DMC96G10	I17281 Sequence 7
40	38.2	6.4	2943	5	I17281	M57388 Simian herp
41	38.2	6.4	2943	16	HSVGBQ	AF126748 Homo sapi
42	38	6.4	1159	40	AF126748	AF126749 Homo sapi
43	38	6.4	1472	40	AF126749	Y08701 M.musculus
44	38	6.4	2813	12	MMPININ	M33782 Human IFEB
45	37.8	6.4	1526	9	HUMTFEB	

ALIGNMENTS

RESULT 1  
AR047920  
LOCUS  
DEFINITION Sequence 1 from patent US 5820864.  
ACCESSION AR047920  
VERSION AR047920.1  
KEYWORDS  
3402 bp DNA  
PAT 29-SEP-1999  
GI:5970263

SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
1 (bases 1 to 3402)  
AUTHORS Paranhos-Baccala, G., Lesenechal, M. and Jolivet, M.  
TITLE Trypanosoma cruzi antigen, gene encoding therefor and methods of  
detecting and treating chagas disease  
JOURNAL Patent: US 5820864-A 1 13-OCT-1998;  
FEATURES Location/Qualifiers  
source  
1. 3402  
BASE COUNT 888 a 821 c 956 g 737 t  
ORIGIN  
Query Match 100.0%; Score 594; DB 5; Length 3402;  
Best Local Similarity 100.0%; Pred. No. 4.4e-117;  
Matches 594; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CAGGTACAGCGTACGGCTTTTGGCTTCAATCGTACAGCCGACGGTAGCTGGCTCGCT 60  
Db 1232 CAGGTACAGCGTACGGCTTTTGGCTTCAATCGTACAGCCGACGGTAGCTGGCTCGCT 1291  
QY 61 GACATGTCGATTGACGATCTCCATCTCCGGTCTCCCGCAGGGAAGAACACAG 120  
Db 1292 GACATGTCGATTGACGATCTCCATCTCCGGTCTCCCGCAGGGAAGAACACAG 1351  
QY 121 CCAGGCCAAAACATCGGTAGTGGCGACGGCGAAACCGGGGTGTGTCTCTCGGGCACT 180  
Db 1352 CCAGGCCAAAACATCGGTAGTGGCGACGGCGAAACCGGGGTGTGTCTCTCGGGCACT 1411  
QY 181 GACGCGGAGTAGCAGTACATACCAATACGACTTCTCCGCTGTGTGATCCCTGCAATCA 240  
Db 1412 GACGCGGAGTAGCAGTACATACCAATACGACTTCTCCGCTGTGTGATCCCTGCAATCA 1471  
QY 241 CCCCTGTTTCAGCGCCAGCAAGCGCGCTCTTCCGCGGCGCATCGCTGAG 300  
Db 1472 CCCCTGTTTCAGCGCCAGCAAGCGCGCTCTTCCGCGGCGCATCGCTGAG 1531  
QY 301 CCGCAGCTGGGGAGCAAGATCATTTGCTAAATCTAGTGAATCAGCTGGGGATTAAATGTCA 360  
Db 1532 CCGCAGCTGGGGAGCAAGATCATTTGCTAAATCTAGTGAATCAGCTGGGGATTAAATGTCA 1591  
QY 361 CAAAGGAGCGTCTCAGCAGTGGAGCGCCGACAGGAGTCTAGCGGGTGACGCTCC 420  
Db 1592 CAAAGGAGCGTCTCAGCAGTGGAGCGCCGACAGGAGTCTAGCGGGTGACGCTCC 1651  
QY 421 ACAGTACCGCCCGCAGCGAACAAGTCCATACGGGCAATGGCCGACCTGTGACGGCT 480  
Db 1652 ACAGTACCGCCCGCAGCGAACAAGTCCATACGGGCAATGGCCGACCTGTGACGGCT 1711  
QY 481 GGATTGTGGCAGTAATAGTGGTCCAGCGCGCTCTGCTCCACAGCGCGCGGCAAA 540  
Db 1712 GGATTGTGGCAGTAATAGTGGTCCAGCGCGCTCTGCTCCACAGCGCGCGGCAAA 1771  
QY 541 CCAACAGGAGAGAAAGGCTCCCGGCGATGTGAACAGAGCTCCGTTGGCGATA 594  
Db 1772 CCAACAGGAGAGAAAGGCTCCCGGCGATGTGAACAGAGCTCCGTTGGCGATA 1825  
RESULT 2  
LOCUS A48910 3402 bp DNA PAT 07-MAR-1997  
DEFINITION Sequence 1 from Patent WO9605312.  
ACCESSION A48910  
VERSION A48910.1 GI:2302570  
KEYWORDS  
SOURCE Trypanosoma cruzi.  
ORGANISM Trypanosoma cruzi  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
Trypanosoma; Schizotrypanum.  
REFERENCE 1 (bases 1 to 3402)  
AUTHORS Paranhos-Baccala, G., Lesenechal, M. and Jolivet, M.

TITLE NOVEL TRYPANOSOMA CRUZI ANTIGEN, AND GENE CODING THEREFOR; THEIR  
APPLICATION TO THE DETECTION OF CHAGAS' DISEASE  
JOURNAL Patent: WO 9605312-A 1 22-FEB-1996;  
COMMENT BIO MERIEUX (FR)  
Other publication CA 2173957 960222  
Other publication AU 3169195 960307  
Other publication FR 2723589 960216.  
FEATURES Location/Qualifiers  
source  
1. 3402  
BASE COUNT 889 a 818 c 958 g 737 t  
ORIGIN  
Query Match 99.5%; Score 590.8; DB 5; Length 3402;  
Best Local Similarity 99.7%; Pred. No. 2.1e-116;  
Matches 592; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CAGGTACAGCGTAAACGGCTTTTGGCTTCAATCGTACAGCCGACGGTAGCTCGCTCGCT 60  
Db 1232 CAGGTACAGCGTAAACGGCTTTTGGCTTCAATCGTACAGCCGACGGTAGCTCGCTCGCT 1291  
QY 61 GACATGTCGATTGATGACGATCTTCCATCTCCGGTCTCCCGCAGGGAAGAACACAG 120  
Db 1292 GACATGTCGATTGATGACGATCTTCCATCTCCGGTCTCCCGCAGGGAAGAACACAG 1351  
QY 121 CCAGGCCAAAACATCGGTAGTGGCGACGGCGAAACCGGGGTGTGTCTCTCGGGCACT 180  
Db 1352 CCAGGCCAAAACATCGGTAGTGGCGACGGCGAAACCGGGGTGTGTCTCTCGGGCACT 1411  
QY 181 GACGCGGAGTAGCAGTACATACCAATACGACTTCTCCGCTGTGTGATCCCTGCAATCA 240  
Db 1412 GACGCGGAGTAGCAGTACATACCAATACGACTTCTCCGCTGTGTGATCCCTGCAATCA 1471  
QY 241 CCCCTGTTTCAGCGCCAGCAAGCGCGCTCTTCCGCGGCGCATCGCTGAG 300  
Db 1472 CCCCTGTTTCAGCGCCAGCAAGCGCGCTCTTCCGCGGCGCATCGCTGAG 1531  
QY 301 CCGCAGCTGGGGAGCAAGATCATTTGCTAAATCTAGTGAATCAGCTGGGGATTAAATGTCA 360  
Db 1532 CCGCAGCTGGGGAGCAAGATCATTTGCTAAATCTAGTGAATCAGCTGGGGATTAAATGTCA 1591  
QY 361 CAAAGGAGCGTCTCAGCAGTGGAGCGCCGACAGGAGTCTAGCGGGTGACGCTCC 420  
Db 1592 CAAAGGAGCGTCTCAGCAGTGGAGCGCCGACAGGAGTCTAGCGGGTGACGCTCC 1651  
QY 421 ACAGTACCGCCCGCAGCGAACAAGTCCATACGGGCAATGGCCGACCTGTGACGGCT 480  
Db 1652 ACAGTACCGCCCGCAGCGAACAAGTCCATACGGGCAATGGCCGACCTGTGACGGCT 1711  
QY 481 GGATTGTGGCAGTAATAGTGGTCCAGCGCGCTCTGCTCCACAGCGCGCGGCAAA 540  
Db 1712 GGATTGTGGCAGTAATAGTGGTCCAGCGCGCTCTGCTCCACAGCGCGCGGCAAA 1771  
QY 541 CCAACAGGAGAGAAAGGCTCCCGGCGATGTGAACAGAGCTCCGTTGGCGATA 594  
Db 1772 CCAACAGGAGAGAAAGGCTCCCGGCGATGTGAACAGAGCTCCGTTGGCGATA 1825  
RESULT 3  
LOCUS TCU24190  
DEFINITION Trypanosoma cruzi 3402 bp mRNA INV 04-AUG-1997  
ACCESSION U24190  
VERSION U24190.1 GI:790645  
KEYWORDS  
SOURCE Trypanosoma cruzi.  
ORGANISM Trypanosoma cruzi  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
Trypanosoma; Schizotrypanum.



binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES	Location/Qualifiers	misc_feature
source	1. .47852 /organism="Mycobacterium tuberculosis" /strain="H37Rv" /db_xref="taxon:1773"	gene
source	<1. .18 /organism="Mycobacterium tuberculosis" /strain="H37Rv" /db_xref="taxon:1773" /clone="Y13E12" complement(1. .1695) /gene="Rv3494c" /note="Rv3494c, (MTV023.01c), len: 564. Unknown Pro-rich protein similar to several Mycobacterium tuberculosis proteins e.g. MTC128.14 (515 aa), MTCY19H5.28c (516 aa) and (MTV051.09). Has hydrophobic stretch, possibly signal peptide at N-terminus. FASTA scores: 297050 MTC128_14 (515 aa) opt: 1006 z-score: 582.8 E(): 5.9e-25; 34.3% identity in 559 aa overlap: 297182 MTCY19H5_28 (516 aa) opt: 979 z-score: 567.7 E(): 4.1e-24; 33.5% identity in 555 aa overlap. TBparse score is 0.897"	CDS
CDS	/transl_table=1 /product="hypothetical protein Rv3494c" /protein_id="CAAL7731.1" /db_xref="GI:2924431" /db_xref="SPTREMBL:O53539" /translation="MDRLAKIOLSFIVITVTLVSMAIFYLRLPATRGITYGVSADFVAGGLYKNANTYRGVAGVSGVGLNPGVTAHRLNSCTAIPSNVATVRSVSAIGEQLIDLPNPNPSTKLRNGFRQORNRIGQVDALLROAELLGSLGTRURELHEAATAGAPGELARLIESARLLVDEANAYPOVSOLIDAGPFLQAIKAGGDIKSLADLRTWLQRAEDPLRDLTADAPDAIDEANTAFSGIRPSFPALAAIANLGRVVIYHKSIEQLVFPALFAALITISAGGVQDEGAKLDFKIDLHPDPCMTGFLFPPVLRVPSADEIREIPDMYCKTAQNDPSTVGRARNYPCOEFGPKRAPTVOLCRDPGCVYPVCTNWRGPPPIPYGVTEVDGRLNLPNKEFPYIPGADDPGVPIGVPPPPQVAGPGPAHQPAQAPNDNGPPPPFSWPGYPPPEPQVPIPTIPPPPPPEGGTGPFGPGAPGQPAAGPATIYDQUSGAFADPAGGTGIFAPGTGASSAENWVLDLRDPRQL" complement(1. .1695) /gene="Rv3494c" /note="Rv3494c" <1. .47852 /note="fragment designated v023. Does not represent a physical clone" complement(1703. .1707) /note="possible RBS for Rv3494c" complement(1706. .2860) /gene="lprN" complement(1706. .2860) /gene="lprN" /note="Rv3495c, (MTV023.02c), len: 384. lprN, similar to Mycobacterium tuberculosis proteins MTC128.13 (390 aa) and MTCY19H5.29 (402 aa) and (MTV051.08). Probably lipoprotein, contains possible signal sequence and appropriately positioned PS00013 Prokaryotic membrane lipoprotein lipid attachment site. FASTA scores: 91 297050 MTC128_13 (390 aa) opt: 653 z-score: 762.1 E(): 0; 33.6% identity in 363 aa overlap; and 297182 MTCY19H5_29 (402 aa) opt: 572 z-score: 667.9 E(): 1.1e-29; 31.8% identity in 362 aa overlap. TBparse score is 0.897" /codon_start=1 /transl_table=1 /product="lprN" /protein_id="CAAL7732.1" /db_xref="GI:2924432" /db_xref="SPTREMBL:O53540" /translation="MNRILWRLAIITASSALLAGCGGINSPLPCTAGGEGAYSVTVENADYALPQNSPVHVDVTVGSVAGIVAVORPGSVYAKVLDLDRNLLPANAVAKVQYSLGSLSHVLPADPTDRPTGLVDGSRITANTDRPTTEVFSALGVVYKGNVGALEETIDETHAVAGRQAVNLVPLRLAELTAGLNKRVHDIIDALDGLNRVSAI LARDKDNLRALDLPDAVRVLRNQNRDHIVDFAALKRLTMTVTSVLAEKTVDFGDEL	gene
misc_feature	<1. .47852 /note="fragment designated v023. Does not represent a physical clone" complement(1703. .1707) /note="possible RBS for Rv3494c" complement(1706. .2860) /gene="lprN" complement(1706. .2860) /gene="lprN" /note="Rv3495c, (MTV023.02c), len: 384. lprN, similar to Mycobacterium tuberculosis proteins MTC128.13 (390 aa) and MTCY19H5.29 (402 aa) and (MTV051.08). Probably lipoprotein, contains possible signal sequence and appropriately positioned PS00013 Prokaryotic membrane lipoprotein lipid attachment site. FASTA scores: 91 297050 MTC128_13 (390 aa) opt: 653 z-score: 762.1 E(): 0; 33.6% identity in 363 aa overlap; and 297182 MTCY19H5_29 (402 aa) opt: 572 z-score: 667.9 E(): 1.1e-29; 31.8% identity in 362 aa overlap. TBparse score is 0.897" /codon_start=1 /transl_table=1 /product="lprN" /protein_id="CAAL7732.1" /db_xref="GI:2924432" /db_xref="SPTREMBL:O53540" /translation="MNRILWRLAIITASSALLAGCGGINSPLPCTAGGEGAYSVTVENADYALPQNSPVHVDVTVGSVAGIVAVORPGSVYAKVLDLDRNLLPANAVAKVQYSLGSLSHVLPADPTDRPTGLVDGSRITANTDRPTTEVFSALGVVYKGNVGALEETIDETHAVAGRQAVNLVPLRLAELTAGLNKRVHDIIDALDGLNRVSAI LARDKDNLRALDLPDAVRVLRNQNRDHIVDFAALKRLTMTVTSVLAEKTVDFGDEL	RBS
gene	complement(1703. .1707) /note="possible RBS for Rv3494c" complement(1706. .2860) /gene="lprN" complement(1706. .2860) /gene="lprN" /note="Rv3495c, (MTV023.02c), len: 384. lprN, similar to Mycobacterium tuberculosis proteins MTC128.13 (390 aa) and MTCY19H5.29 (402 aa) and (MTV051.08). Probably lipoprotein, contains possible signal sequence and appropriately positioned PS00013 Prokaryotic membrane lipoprotein lipid attachment site. FASTA scores: 91 297050 MTC128_13 (390 aa) opt: 653 z-score: 762.1 E(): 0; 33.6% identity in 363 aa overlap; and 297182 MTCY19H5_29 (402 aa) opt: 572 z-score: 667.9 E(): 1.1e-29; 31.8% identity in 362 aa overlap. TBparse score is 0.897"	gene
CDS	/transl_table=1 /product="lprN" /protein_id="CAAL7732.1" /db_xref="GI:2924432" /db_xref="SPTREMBL:O53540" /translation="MNRILWRLAIITASSALLAGCGGINSPLPCTAGGEGAYSVTVENADYALPQNSPVHVDVTVGSVAGIVAVORPGSVYAKVLDLDRNLLPANAVAKVQYSLGSLSHVLPADPTDRPTGLVDGSRITANTDRPTTEVFSALGVVYKGNVGALEETIDETHAVAGRQAVNLVPLRLAELTAGLNKRVHDIIDALDGLNRVSAI LARDKDNLRALDLPDAVRVLRNQNRDHIVDFAALKRLTMTVTSVLAEKTVDFGDEL	CDS

KDLYSTIVKALNDRKDFVTSLOLLTFPPFPNFGIKQAVRGDYLNVFTTDLTLRRIGE TFFTTAYFDPNNAHDEILNPPDFLIGELANLSGQADDFKIPPGTASQ"  
complement(2798. .2830)  
/gene="lprN"  
/note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"  
complement(2857. .4212)  
/gene="Rv3496c"  
complement(2857. .4212)  
/gene="Rv3496c"  
/note="Rv3496c, (MTV023.03c), len: 451. Unknown but similar to Mycobacterium tuberculosis proteins MTC128.12 (530 aa), MTCY19H5.30c (508 aa) (MTV051.07). Hydrophobic region at N-terminus. FASTA scores: 297050|MTC128\_12 (530 aa) opt: 838 z-score: 711.0 E(): 4.3e-32; 33.1% identity in 473 aa overlap; and 297182|MTCY19H5\_30 (508 aa) opt: 821 z-score: 697.0 E(): 2.6e-31; 35.1% identity in 453 aa overlap. TBparse score is 0.891"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein Rv3496c"  
/protein\_id="CAAL7733.1"  
/db\_xref="GI:2924433"  
/db\_xref="SPTREMBL:O53541"  
/translation="MMRVAMLTGSRLEAYATVIALVAALGVVYVLSGTGKRTIVGYTSAVGLYPGQVRVLGVPVGEIDMIERSDDVKITMSVSKDKVPDQVIMSPN LVAARFQILPTVITGAVLPDNGRIDLRTAVPVEWDEKGLTAAALSPAAGELQ GPALAAINQADTLIDGNGSLHNAELRAQVAGRLGDSRGDIFGTNQLQVLDALSE SDEQIVQFAGHAVSVQSLVADSANLDOTLGTNLQALSDIRGLFRENNSLTLETNQL NDFQOTLDSQSENIQOVLHVAGPITNFYNIYDPAQGLNLGLSLINFPANPQVFCGG SFTDAAGSPADPYRRAETICRELGLPRLVNPVIMFHLNITATKVGQIIVDTP ATEAKSETPVPLTWVAGGAPVGNADQLSLVPPAPGAPAPGAPGAGEHGGG G"  
complement(2867. .2871)  
/gene="Rv3496c"  
/note="possible RBS for Rv3495c"  
complement(4209. .5282)  
/gene="Rv3497c"  
complement(4209. .5282)  
/gene="Rv3497c"  
/note="Rv3497c, (MTV023.04c), len: 357. Unknown but similar to Mycobacterium tuberculosis proteins MTCY19H5.31 (481 aa), MTC128.11 (515 aa) and MTV051.06. Hydrophobic region at N-terminus. FASTA scores: 297182|MTCY19H5\_31 (481 aa) opt: 611 z-score: 667.7 E(): 1.1e-29; 32.5% identity in 332aa overlap; and 297050|MTC128\_11 (515 aa) opt: 587 z-score: 641.2 E(): 3.3e-28; 30.1% identity in 335 aa overlap. TBparse score is 0.889"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein Rv3497c"  
/protein\_id="CAAL7734.1"  
/db\_xref="GI:2924434"  
/db\_xref="SPTREMBL:O53542"  
/translation="MLNRKPSKHERDLRTGIFGLVLVICVLIATFGSLPFWPQG KYIDAFDTAGITFGNSVYVSGLKVGAVSAVSLGNSAKVIFSDVSLVWGDSLAA IRTDITLGRSIAVSPAGSKSTIPLSTTTPTTLNGVLQDLGRNANDLNRPQFQA LNVFTQALHDATPQVRGAVDGLTSLNRDRDEALQGLLAHSAVTSYLSERAQVN KIVEDGNOLFALDARRAALSALISIDIDVAAQISGFVADNRKEFGPALSRLNLVAN LNERDYITPEALKRLPTATTILGEVVGSGPFGFNVYSLVPLGVLATVDFLVFPQGL PSLADYLRGFIQERWILRPKSP"  
complement(5272. .5324)  
/gene="Rv3498c"  
complement(5272. .6324)  
/gene="Rv3498c"  
/note="Rv3498c, (MTV023.05c), len: 350. Unknown but similar to Mycobacterium tuberculosis proteins MTC128.10 (346 aa), MTCY19H5.32c (346 aa) and MTV051.05. Hydrophobic region at N-terminus. FASTA scores: 297050|MTC128\_10 (346 aa) opt: 755 z-score: 883.6 E(): 0; 36.8% identity in 340 aa overlap; and 297182|MTCY19H5\_32 Mycobacterium tuberculosis cosmi (275 aa) opt: 699z-score: 820.1 E(): 0; 47.0% identity in 249 aa overlap. TBparse score is 0.878"

```
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3498c"
/protein_id="CAA17735.1"
/db_xref="GI:2924435"
/db_xref="SPTREMBL:O53543"

Query Match      7.6%; Score 45; DB 1; Length 47852;
Best Local Similarity 43.7%; Pred. No. 1.9; Mismatches 255; Indels 0; Gaps 0;
Matches 198; Conservative 0;

Qy 118 CAGCCAGGCCAAACATCGTAGTCGCGAGCGGGAACCGGGGTGTGTCTCTCGGGC 177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 32502 CCGCGGGCAAGCGCGCGCGCATGGCGCAACGGGSCCATCGGGCCACGCGCC 32561

Qy 178 ACTGACGGCGGAGTAGCAGTCATACCAATACGACTTCTGCGGTGCTGATCCCTGCA 237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 32562 CACTCACTGACGACCGCCCGCGCAACGGGGCACCGGGCAACGGCGCACCGCGGCA 32621

Qy 238 TCACCCCTGTTTCAGCCGACGCAAGCAGCGCGCTCTCTGCGCGCGCGGATCGGCT 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 32622 CCGCGGGCGGGCATCGGACCTTGGCGGCGCATGCGGCGCATGGCGCAACGCGC 32681

Qy 298 GAGCCGACGTGGGAGCAATGCTATCTAGTGAATCAGCTGGGATTAATGTC 357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 32682 GCAACGGCGGTACCGCGGAGCGCGAGGTGCGGCGCGCGCGCACCGCGCGTG 32741

Qy 358 ACCCAAGAGGCGTCTGACGACTGGAGCGCGCGCACGACGAGTCTACGGCGGTGACG 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 32742 CGGCGGCAATGCGGCGCATGGCGCACCGCGCGCGCGCGCGCGCGCGCGCG 32801

Qy 418 TCACGACTACCGCGCGCGCGCAAGTCCATACGCGCAATGCGGCGCTGTGACG 477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 32802 GCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 32861

Qy 478 GCTGATTGTTGAGCTAATAGTGTGCGAGCGGGCTCTGCTCTCCACAGCGCGCGG 537
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 32862 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 32921

Qy 538 AAACAACAGGAGAAAGCGCTCTCGGGGCA 570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 32922 ACGCGCGCGGAGGCAATGCAACGCGGCA 32954

RESULT 5
MLURHO
LOCUS MLURHO 2402 bp DNA BCT 12-NOV-1996
DEFINITION Micrococcus luteus Rho factor (rho) gene, complete cds.
ACCESSION L27277
VERSION L27277.1 GI:1666539
KEYWORDS Micrococcus luteus.
SOURCE Micrococcus luteus.
ORGANISM Micrococcus luteus
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Micrococciaceae; Micrococcaceae; Micrococcus.
1 (bases 905 to 2402)
Opperman, T. and Richardson, J.P.
Phylogenetic analysis of sequences from diverse bacteria with
homology to the Escherichia coli rho gene
J. Bacteriol. 176 (16), 5033-5043 (1994)
94327472
2 (bases 1 to 2402)
Nowatzke, W.L. and Richardson, J.P.
Characterization of an unusual Rho factor from the high G + C
gram-positive bacterium Micrococcus luteus
J. Biol. Chem. 271 (2), 742-747 (1996)
96132802
3 (bases 905 to 2402)
Richardson, J.P.
Direct Submission
Submitted (14-FEB-1994) John P. Richardson, Chemistry, Indiana
University, Bloomington, IN 47405, USA
4 (bases 1 to 2402)
```

```
AUTHORS Nowatzke, W.L.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-1996) William L. Nowatzke, Chemistry, Indiana
University, Bloomington, IN 47405, USA
COMMENT On Nov 12, 1996 this sequence version replaced gi:968908.
FEATURES
    Location/Qualifiers
        1..2402
            /organism="Micrococcus luteus"
            /strain="EM"
            /db_xref="taxon:1270"
            288..2363
            /gene="rho"
            288..2363
            /gene="rho"
            /function="transcription termination factor"
            /codon_start=1
            /transl_table=11
            /product="Rho factor"
            /protein_id="AA018671.1"
            /db_xref="GI:1666540"
            /translation="MTSTQTTPNGGLASLKLQALQALASOLGAGGSRMRKADL
            VTAISDHORGSSVADRAAERAAQAPAAETAPAAASSEDAAAPAAERARRSRA
            DADTSAPAAQDQGPQAEAREAQTEAPRETASQDRSGGSEADEGDRPQSSRRS
            GRRAGDDDAQOGDRSDGAQGEDGADADRDRDDNGRNGRNGRNGRNGRNGR
            NGRDRENGRNSRDRENGRSGRSGRSGRSGRSGRSGRSGRSGRSGRSGR
            NRRNRRNGRNRGRGPEVDELTEDDVLQPVAGILDVLDNTAFVTSGLPGPND
            VYSLAMVKYGLRKDAVGPPIAPRDEKQHHGGSNRKFNALKRISVNGOPAV
            EHPQRFVFKLPLVLPQERLETPDKLIGRPVLDLSPGKGRLIVSPKAGTM
            ILASIANIKTNPEVHLMMVLVDEPEREVTDMQSRVDSGEVIAGTFDRPADDTLAE
            LAIRAKRLVMGRDVLVLLDSTMTLRGLRAYNLAAAPSGRLSGVDSALYPPKFFG
            AARNIENGSSITILATLVETGSRMDEVFEEFKGTGTMELRLSRHLAERFIQAVDV
            NASGTREALLSQQEYKIMWKLRRVLSGLEQQALDLITNKIRDTASNAEFLMLVSK
            TTUGSKGDD"
BASE COUNT 386 a 818 g 251 t
ORIGIN

Query Match      7.4%; Score 44; DB 1; Length 2402;
Best Local Similarity 45.6%; Pred. No. 4;
Matches 113; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

Qy 367 AGCGTGGTCAGCACTGGAGCGCGCCACGACGAGGTCTACGGCGGTGACGTCACGACT 426
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 799 AGCAGGGTCAGGACCGTCTCCGACGCGGCCCGGAGGACGCGCCGCGGACCC 858

Qy 427 ACCGCGCGCGGCAACAAAGTCCATACGGGCGACAAATGGCCGACCTGTGACGCTGATTG 486
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 859 GCGCGCGCGCGGAGGACCGGTGACGACAAAGCGGCGCGGAGAACGCGCGCGCAACG 918

Qy 487 GTGGCAGCTAATAGTGTGGCCAGCGCGGCGCTGTCTCCACAGCGCGCGGCAACCA 546
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 919 GCGGCAACGCGCGGAGGACCGCGCAACGCGCGGCGGAGAACGCGCGCGGAGAACG 978

Qy 547 GGAGAGAAAGGCGCTGGGCGCATGTGAACGAGCTCGTGGGAGATA 594
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 979 GGGACCGTGAGAACGCGCGCGGCGCGGCGCTCGCGGAGCAGCGCGCGGAGAACG 1026

RESULT 6
SS1132828/c
LOCUS SS1132828 2028 bp mRNA PLN 05-APR-1999
DEFINITION Spmatozopsis similis mRNA for p210 protein, partial.
ACCESSION AJ132828
VERSION AJ132828.1 GI:4584085
KEYWORDS p210 protein.
SOURCE Spmatozopsis similis.
ORGANISM Spmatozopsis similis
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
Spmatozopsis.
1 (bases 1 to 2028)
Lechtreck, K.F., Teltienkoetter, A. and Grunow, A.
A 210-kDa-protein is located in a membrane-microtubule-linker at
the distal end of mature and nascent basal bodies
```

[illegible]













```

/note="Rv3348, (MTV004.04), len: 163 aa; unknown,
identical to M. tuberculosis hypothetical protein
TR:P96234 (EMBL:783864) MTCV01A6.25 (163 aa). Partially
similar to several insertion elements, e.g.
P198341Y111_STRCL_INSERTION_ELEMENTIS116, (399 aa, 28.5%
identity in 158 aa overlap).",
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv3348"
/protein_id="CAAL5733.1"
/db_xref="GI:2661627"
/db_xref="SPTREMBL:P96234"
/transl_table="MTAENPGRSRTLVGIDAAITACHIAIRDVGCARSIRFSVEPT
LAGRLTDLKLSGVDDIDATVEPTSMTWLPTIAVENAGDTMHAGARHCAIRGALV
LAKSRSDVIDAEVLTRASEVFDLTPLTPTPAQLALRSVIRRAGAVIDANRSRLMS
LAR"
complement(11239..11979)
/gene="Rv3349c"
/cds
complement(11239..11979)
/gene="Rv3349c"
/note="Rv3349c, (MTV004.05c), len: 246 probable
transposase pseudogene fragment, similar to part of
TR:Q50911 (EMBL:U10634) IS204 PUTATIVE TRANSPOSASE from
NOCARDIA ASTEROIDES (377 aa), fasta scores; opt: 288
z-score: 322.6 E(): 8.3e-11, 48.5% identity in 97 aa
overlap"
/codon_start=1

Query Match
Best Local Similarity 6.8%; Score 40.4; DB 1; Length 69350;
Matches 86; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 142 GTGGCGACGGCGAACCAGGGGTGTCTCTGGGCATGTAGCGCGGAGTAGCAGTCAAT 201
Db 37024 GCGGCGACGGCGGTGACGGGGGACGCGGACCGCGGTGACGGGTACTGGCGG 37083

QY 202 ACCAATACGACTTCGCGCTGCTCATCCCTGCATCATCCCGTTCAGCGCCAGCC 261
Db 37084 GCACACGCGCAACTTCGGCAATGCGGTGGCGCGGTGACCGCGGCGGCGCC 37143

QY 262 AAGCGACGGCCCTCTCTGCGCGCGGCGATGCGGTGAGCGCC 303
Db 37144 ATGGCACCCCGGACGCGCAACGCGGCGGCGCACCGCGCACCG 37185

RESULT 14
AP000556/c
LOCUS
DEFINITION Homo sapiens genomic DNA, chromosome 22q11.2, BCRL2 region,
clone:KB1172D5.
ACCESSION AP000556
VERSION AP000556.2 GI:6139049
KEYWORDS
SOURCE Homo sapiens DNA, clone:KB1172D5.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 149618)
Shimizu, N.
Human DNA sequence from clone KB1172D5 on chromosome 22q11.2
Published Only in Database (1999) In press
2 (bases 1 to 149618)
Shimizu, N.
Direct Submission
Submitted (01-OCT-1999) to the DDBJ/EMBL/GenBank databases.
Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular
Biology, 35 Shinanouchi, Shinjuku-ku, Tokyo 160-0015, Japan
(E-mail:nshimizu@db.med.keio.ac.jp, Tel:81-3-3351-2370,
Fax:81-3-3351-2370)
On Oct 29, 1999 this sequence version replaced gi:6006352.
This is a complete sequence of the insert of KB1172D5 clone. The
proximal adjacent clone is KB1183D5 (Acc.#AP000552) with 67422-bp
overlapping. The distal adjacent clone is KB1323B2 (Acc.#AP000557)

```

```

with 69793-bp overlapping
Sequence updated (26-Oct-1999).
FEATURES
source 1. 149618
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="KB1172D5"
/map="22q11.2"
complement(1..667)
/rpt_family="L1M4c"
/evidence-not_experimental
complement(1010..1716)
/rpt_family="L1M4c"
/evidence-not_experimental
complement(1717..2021)
/rpt_family="AluSx"
/evidence-not_experimental
complement(2022..2297)
/rpt_family="L1M4c"
/evidence-not_experimental
2297..2368
/rpt_family="L1MB4"
/evidence-not_experimental
2760..3043
/rpt_family="AluSq"
/evidence-not_experimental
3044..3066
/rpt_family="(TAAA)n"
/evidence-not_experimental
complement(4122..4305)
/rpt_family="AluSg/x"
/evidence-not_experimental
4949..5151
/rpt_family="AluJb"
/evidence-not_experimental
5649..5748
/rpt_family="MIR"
/evidence-not_experimental
complement(8880..9173)
/rpt_family="AluY"
/evidence-not_experimental
10813..10860
/rpt_family="GC-rich"
/evidence-not_experimental
10862..10892
/rpt_family="(CGG)n"
/evidence-not_experimental
11091..11118
/rpt_family="GC-rich"
/evidence-not_experimental
complement(11968..12299)
/rpt_family="AluJo"
/evidence-not_experimental
13437..13529
/rpt_family="MIR"
/evidence-not_experimental
15901..16312
/rpt_family="MER31B"
/evidence-not_experimental
16308..16476
/rpt_family="MER31B"
/evidence-not_experimental
complement(16477..16655)
/rpt_family="AluSc"
/evidence-not_experimental
17091..17265
/rpt_family="L2"
/evidence-not_experimental
17219..17272
/rpt_family="MIR"
/evidence-not_experimental
17597..17671

```

```
repeat_region /rpt_family="G-rich"
/evidence-not_experimental
complement(17672..17831)
/rpt_family="AluSx"
repeat_region /evidence-not_experimental
complement(17832..17880)
/rpt_family="MER45C"
repeat_region /evidence-not_experimental
21993..22302
/rpt_family="AluY"
repeat_region /evidence-not_experimental
22304..22737
/rpt_family="L1MD3"
repeat_region /evidence-not_experimental
22888..23013
/rpt_family="AluYb8"
repeat_region /evidence-not_experimental
23014..23323
/rpt_family="AluY"
repeat_region /evidence-not_experimental
23324..23496
/rpt_family="AluYb8"
repeat_region /evidence-not_experimental
complement(24098..24406)
/rpt_family="AluYb"
repeat_region /evidence-not_experimental
complement(24447..24558)
/rpt_family="MIR"
repeat_region /evidence-not_experimental
24631..24721
/rpt_family="CT-rich"
repeat_region /evidence-not_experimental
24952..25245
/rpt_family="AluSx"
repeat_region /evidence-not_experimental
26011..26140
/rpt_family="L2"
repeat_region /evidence-not_experimental
26310..26339
/rpt_family="(TTTG)n"
repeat_region /evidence-not_experimental
complement(26340..26623)
/rpt_family="AluSx"
repeat_region /evidence-not_experimental
26731..26833
/rpt_family="L2"
repeat_region /evidence-not_experimental
complement(26798..26898)
/rpt_family="L2"
repeat_region /evidence-not_experimental
26899..27029
/rpt_family="FLAM_C"
repeat_region /evidence-not_experimental
complement(27037..27337)
/rpt_family="AluY"
repeat_region /evidence-not_experimental
complement(27338..27380)
/rpt_family="L2"
repeat_region /evidence-not_experimental
27868..28160
/rpt_family="AluY"
repeat_region /evidence-not_experimental
28255..28286
/rpt_family="(TTTTA)n"
repeat_region /evidence-not_experimental
complement(28298..28602)
/rpt_family="AluYb"
repeat_region /evidence-not_experimental
complement(28612..28912)
/rpt_family="AluSx"
repeat_region /evidence-not_experimental
complement(28927..29192)
/rpt_family="AluSg"
```

```
repeat_region /evidence-not_experimental
complement(29229..29542)
/rpt_family="AluSx"
repeat_region /evidence-not_experimental
complement(29544..29741)
/rpt_family="7SLRNA"
repeat_region /evidence-not_experimental
29808..30107
/rpt_family="AluY"
repeat_region /evidence-not_experimental
complement(30206..30372)
/rpt_family="L2"
repeat_region /evidence-not_experimental
complement(30521..30840)
/rpt_family="AluY"
repeat_region /evidence-not_experimental
complement(30841..30969)
/rpt_family="AluYb"
repeat_region /evidence-not_experimental
complement(31012..31228)
/rpt_family="AluJo"
repeat_region /evidence-not_experimental
31307..31387
/rpt_family="MIR"
repeat_region /evidence-not_experimental
31540..31819
/rpt_family="AluSc"
/evidence-not_experimental

Query Match 6.8%; Score 40.4; DB 10; Length 149618;
Best Local Similarity 52.9%; Pred. No. 17;
Matches 109; Conservative 0; Mismatches 96; Indels 1; Gaps 1;

QY 141 AGTGGCAGCGGAAACCGGGGTGTGTCTCGGGCACTGACGCCGCGAGTAGCAGTCA 200
|| ||| ||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35731 AGCAGCGCGCTGCAGCTGTGTCTCGGCTCTCTGCCACGTCGCGCGCGCGCGCC 35672

QY 201 TACCAATAGACTTTCGCGCTGCTGCATCCCTGCATCACCCCTGTTTCAGCGCAGC 260
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35671 TCGTGCCGAGCTCTCGGCATCGCGCGCGCTGCTGCGCTCGCCAGCGCCAGC 35612

QY 261 CAAGGAGCGCGCGCTCTCGCGCGCGGATC-GGCTGAGCGGCACGTCGGGAGCAAGA 319
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35611 ATCTGCGCTCGCACCTCTCTCGAGGTTGCGCTCTGCGCGCGCGCTCGGCCAAGCCCA 35552

QY 320 TCATTGCTAATCTAGTGAATCAGCTG 345
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35551 CGTTGCTGCTGAAGCATCACTG 35526

RESULT 15
AP000557
LOCUS AP000557 150036 bp DNA PRI 02-NOV-1999
DEFINITION Homo sapiens genomic DNA, chromosome 22q11.2, BCRL2 region,
clone:KB1323B2.
ACCESSION AP000557
VERSION AP000557.2 GI:6174873
KEYWORDS
SOURCE Homo sapiens DNA, clone:KB1323B2.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 150036)
AUTHORS Shimizu,N.
TITLE Homo sapiens genomic DNA, chromosome 22q11.2, BCRL2 region, clone
KB1323B2
JOURNAL Published Only in DataBase (1999) In press
AUTHORS Shimizu,N.
TITLE Direct Submission
JOURNAL Submitted (01-OCT-1999) to the DBJ/EMBL/GenBank databases.
Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular
Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan
```

(E-mail:nshimizudmb.med.keio.ac.jp, Tel:81-3-3351-2370;  
Fax:81-3-3351-2370)  
On Nov 2, 1999 this sequence version replaced gi:6006353.  
This is a complete sequence of the insert of KB1323B2 clone. The  
proximal adjacent clone is KB1172D5 (Acc.#AP000556) with 69794-bp  
overlapping. The distal adjacent clone is KB1802C5 (Acc.#AP000558)  
with 21623-bp overlapping.  
Sequence updated (29-Oct-1999).

## COMMENT

## FEATURES

Source	Location/Qualifiers
repeat_region	1..150036 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="22" /clone="KB1323B2" /map="22q11.2" 519..594
repeat_region	/rpt_family="(TTG)n" /evidence=not_experimental complement(554..565)
repeat_region	/rpt_family="FLAM_C" /evidence=not_experimental 566..594
repeat_region	/rpt_family="(TTG)n" /evidence=not_experimental complement(595..716)
repeat_region	/rpt_family="FLAM_C" /evidence=not_experimental 813..1106
repeat_region	/rpt_family="AluSx" /evidence=not_experimental 1918..2204
repeat_region	/rpt_family="AluJb" /evidence=not_experimental 2613..2915
repeat_region	/rpt_family="AluSp" /evidence=not_experimental complement(3253..3353)
repeat_region	/rpt_family="L1ME3A" /evidence=not_experimental complement(3396..3501)
repeat_region	/rpt_family="L1ME3A" /evidence=not_experimental complement(3779..4072)
repeat_region	/rpt_family="AluSx" /evidence=not_experimental 4132..4429
repeat_region	/rpt_family="AluSx" /evidence=not_experimental complement(5553..5635)
repeat_region	/rpt_family="MER5A" /evidence=not_experimental complement(5930..6220)
repeat_region	/rpt_family="AluY" /evidence=not_experimental 7330..7497
repeat_region	/rpt_family="MIR" /evidence=not_experimental complement(9725..9855)
repeat_region	/rpt_family="L1ME3A" /evidence=not_experimental 9901..10191
repeat_region	/rpt_family="AluYb8" /evidence=not_experimental 17988..18048
repeat_region	/rpt_family="(TC)n" /evidence=not_experimental complement(22321..22478)
repeat_region	/rpt_family="HAL1" /evidence=not_experimental 22670..22733
repeat_region	/rpt_family="(TA)n" /evidence=not_experimental complement(22743..22997)
repeat_region	/rpt_family="AluSx" /evidence=not_experimental complement(22998..23084)
repeat_region	/rpt_family="MER58A" /evidence=not_experimental 23091..23406
repeat_region	/rpt_family="AluSx" /evidence=not_experimental 23407..23550
repeat_region	/rpt_family="AluSx" /evidence=not_experimental 23612..23757
repeat_region	/rpt_family="AluSx" /evidence=not_experimental 23758..24046
repeat_region	/rpt_family="AluY" /evidence=not_experimental 24047..24058
repeat_region	/rpt_family="AluSx" /evidence=not_experimental 24059..24357
repeat_region	/rpt_family="AluSx" /evidence=not_experimental 24358..24506
repeat_region	/rpt_family="AluSx" /evidence=not_experimental complement(24507..24815)
repeat_region	/rpt_family="AluY" /evidence=not_experimental 24816..24846
repeat_region	/rpt_family="AluSx" /evidence=not_experimental complement(25153..25450)
repeat_region	/rpt_family="AluJb" /evidence=not_experimental 26278..26587
repeat_region	/rpt_family="AluSx" /evidence=not_experimental complement(26842..26972)
repeat_region	/rpt_family="L2" /evidence=not_experimental 26985..27136
repeat_region	/rpt_family="AluJb" /evidence=not_experimental complement(27163..27322)
repeat_region	/rpt_family="L2" /evidence=not_experimental complement(27951..28039)
repeat_region	/rpt_family="L2" /evidence=not_experimental complement(28041..28351)
repeat_region	/rpt_family="AluY" /evidence=not_experimental 28352..28383
repeat_region	/rpt_family="(TTG)n" /evidence=not_experimental complement(28384..28664)
repeat_region	/rpt_family="AluJb" /evidence=not_experimental complement(28669..28964)
repeat_region	/rpt_family="AluJo" /evidence=not_experimental complement(28968..29280)
repeat_region	/rpt_family="AluJb" /evidence=not_experimental 29359..29484
repeat_region	/rpt_family="AluSx" /evidence=not_experimental complement(29487..29566)
repeat_region	/rpt_family="AluY" /evidence=not_experimental 29606..29747
repeat_region	/rpt_family="FLAM_C"

```
/evidence-not_experimental
29750. .30039
/rpt_family="AluSq"
/evidence-not_experimental
30042. .30356
/rpt_family="AluJo"
/evidence-not_experimental
complement(30945. .31067)
/rpt_family="MLTIC"
/evidence-not_experimental
complement(31068. .31371)
/rpt_family="AluYb8"
/evidence-not_experimental
complement(31372. .31489)
/rpt_family="MLTIC"
/evidence-not_experimental
complement(31508. .31792)
/rpt_family="AluJb"
/evidence-not_experimental
31992. .32012
/rpt_family="Alu_rich"
/evidence-not_experimental
complement(32013. .32138)
/rpt_family="FLAM_C"
/evidence-not_experimental
32185. .32279
/rpt_family="L1M4"
/evidence-not_experimental
32828. .32967
/rpt_family="AluJb"
/evidence-not_experimental
complement(32986. .33124)
/rpt_family="MLTIIH"

Query Match          6.8%; Score 40.4; DB 10; Length 150036;
Best Local Similarity 52.9%; Pred. No. 17;
Matches 109; Conservative 0; Mismatches 96; Indels 1; Gaps 1;

QY 141 AGTGGCAGCGGGAACGGGGGTGTGTCTCGGGCACTGACGGCGGAGTAGCAGTCA 200
|| ||| || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 119367 AGCAGCGCGCGCTGCAGCTGTGTCTCGGCTCTCTCGCCACGCTCGCGCGCGCGCC 119426
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 201 TACCAATACGACTTCTGCCGCTGTGCATCCCTGCATCACCCCTGTTTCAGCGCCAGC 260
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 119427 TGCCTGCCAGCTCCTGGCATTCCGGCGCGCGCTGGTCCAGCTCGCGCTCCAGCGCCAGC 119486
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 261 CAAGGCAGCCCGCCTCTCTCGCGCGCGCGGATC-GGCTGAGCCGACGTGGGGAGCAAGA 319
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119487 ATCTGCCGTGCGACCTCTCTCGCAGTTTGGCTCTGGCGCGCGCGCTCGGGCCAAAGCGCCA 119546
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 320 TCATTGCTAATCTAGTGAATCAGCTG 345
| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 119547 CCGTTGCCCTGCTGAAGCATCACTG 119572
```

Search completed: April 2, 2000, 21:42:57  
Job time: 10416 sec





```

Db 1232 CAGGTACAGCGTACGGCTTTTGGCTTCATCGTACCGGAGCTAGCTGGTCTGGCT 1291
QY 61 GACATGTCGATTCGATTGACGATCTCCATCTCCGGTCTCCCGCAGGGAAGAACAGCAG 120
Db 1292 GACATGTCGATTCGATTGACGATCTCCATCTCCGGTCTCCCGCAGGGAAGAACAGCAG 1351
QY 121 CCAGGCCAAAACATCGGTAGTGGCGGAGCGGGAACCGGGGTGTGTCTCGGGCACT 180
Db 1352 CCAGGCCAAAACATCGGTAGTGGCGGAGCGGGAACCGGGGTGTGTCTCGGGCACT 1411
QY 181 GAGCGCGGAGTAGCAGTCATACCAATACGACTTCTGCCGTGTGTGTCATCCCTGTCATCA 240
Db 1412 GAGCGCGGAGTAGCAGTCATACCAATACGACTTCTGCCGTGTGTGTCATCCCTGTCATCA 1471
QY 241 CCCCTCTTTTACGGCCAGCAAGCGAGCGCGCTCTGCCCGCGCGCATCGGCTGAG 300
Db 1472 CCCCTCTTTTACGGCCAGCAAGCGAGCGCGCTCTGCCCGCGCGCATCGGCTGAG 1531
QY 301 CCGCAGCTGGGAGCAGATCATTCGCTAATCTAGTGAATCAGCTGGGATTAATGTCAAC 360
Db 1532 CCGCAGCTGGGAGCAGATCATTCGCTAATCTAGTGAATCAGCTGGGATTAATGTCAAC 1591
QY 361 CAAGAGGAGCGTCTGACGACTGGAGCGCGGCCACGACGAGGTCTACGGCGGTGACGCTCC 420
Db 1592 CAAGAGGAGCGTCTGACGACTGGAGCGCGGCCACGACGAGGTCTACGGCGGTGACGCTCC 1651
QY 421 ACGACTACGCGCCCGCAGCGAAGTCCATACAGGGGACCAATGCGCGACCTGTGACGGCT 480
Db 1652 ACGACTACGCGCCCGCAGCGAAGTCCATACAGGGGACCAATGCGCGACCTGTGACGGCT 1711
QY 481 GGATTGTGCGCAGTAATAGTGGTCCAGCGCGGCTCTGCCACAGCGCGGGCGAAA 540
Db 1712 GGATTGTGCGCAGTAATAGTGGTCCAGCGCGGCTCTGCCACAGCGCGGGCGAAA 1771
QY 541 CCAACAGGAGAGAAAAGGCTCCCGGCGATGTGAAAACGAGTCCGTGGCGGATA 594
Db 1772 CCAACAGGAGAGAAAAGGCTCCCGGCGATGTGAAAACGAGTCCGTGGCGGATA 1825

RESULT 2
US-08-718-661-1
: Sequence 1, Application US/08718661
: Patent No. 5876972
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: Nucleic acid molecules coding for mammalian
: NUMBER OF SEQUENCES: 15
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/718,661
: FILING DATE:
: CLASSIFICATION: 530
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2790 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 542..2545
US-08-718-661-1

Query Match 6.6%; Score 39.2; DB 3; Length 2790;
Best Local Similarity 54.1%; Pred. No. 0.059;

```

```

Matches 80; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 137 CGGTAGTGGGAGGGGAAACCGGGGTGTGTCTCGGGCACTGACGGCGGAGTAGCA 196
Db 2114 CAGCAGCAGCCACCGCACTGCGCGCGCCACCACCGTGCCTGCGCACAGCCACTGCCA 2173
QY 197 GTCATACCAATAGACTTCTGCGGCTGCTGCATCCCTGCATCCCCCTGTTTCAGCGC 256
Db 2174 GTCCACAGCCATACACAGCCACAGATGCAGCCAGTTTCAGTTGCAGATCCAGCCC 2233
QY 257 CAGCCAAGCAGCGCGGCTCTCTGCCGC 284
Db 2234 CAGATCAGCTACCAAGCTGTGCCGC 2261

RESULT 3
US-08-042-747A-7/c
: Sequence 7, Application US/08042747A
: Patent No. 5487969
: GENERAL INFORMATION:
: APPLICANT: Eberle, Richard
: APPLICANT: Black, Darla
: APPLICANT: Scinicariello, Franco
: APPLICANT: Hilliard, Julia K.
: TITLE OF INVENTION: Cloning and Amplification of Monkey B
: TITLE OF INVENTION: Virus Genes
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cox & Smith Incorporated
: STREET: 112 East Pecan Street, Suite 2000
: CITY: San Antonio
: STATE: Texas
: COUNTRY: USA
: ZIP: 78205
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/042,747A
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Haymond, W. Bradley
: REGISTRATION NUMBER: 35186
: REFERENCE/DOCKET NUMBER: S-0072.179
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 210-554-5500
: TELEFAX: 210-226-8395
: TELEX: 767609
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2943 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 87..2744
US-08-042-747A-7

Query Match 6.4%; Score 38.2; DB 1; Length 2943;
Best Local Similarity 47.7%; Pred. No. 0.11;
Matches 112; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
QY 209 CGACTTCTGCCGTGCTGTCATCCCTGTCATACCCCTGTTTCAGGCGCCAGCAAGCAG 268
Db 255 CGTCTCTGTCGGGAGCTCTCTGCTCGGGAAGCGGGCGGTGGCGGCGGTCGCGCG 196
QY 269 CCGCGGCTCTCTGCGGCGGCGGATCGGCTGAGCGGCACGTGGGGAGCAGATCATTTCTA 328

```



Best local similarity 50.5%, Freq. NO: 0.21,  
Matches 63: Conservative 0: Mismatches 44: Indels 0: Gaps 0:

Qy	258	AGCAAGGCAGCGCGGCTCCTGCCGGCGGATCGGCTGAGCCGC	304
Db	380	ACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCACCAGCTTCTCCAC	426

## RESULT 8













**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: April 2, 2000, 19:13:05 ; Search time 1279.51 seconds  
(without alignments)  
1752.815 Million cell updates/sec

Title: US-08-988-242-1-copy\_1232\_1825  
Perfect score: 594  
Sequence: 1 CAGGTACACGTAACGGCTT.....AAACGAGCTCCGTGGCGATA 594

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues  
Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0  
Maximum DB seq length: 1000000  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :		EST:*	
		1: em_est1.*	45: gb_est26.*
		2: em_est2.*	46: gb_est27.*
		3: em_est3.*	47: gb_est28.*
		4: em_est4.*	48: gb_est29.*
		5: em_est5.*	49: gb_est30.*
		6: em_est6.*	50: gb_est31.*
		7: em_est7.*	51: gb_est32.*
		8: em_est8.*	52: em_est20.*
		9: em_est9.*	53: em_est21.*
		10: em_est10.*	54: em_est22.*
		11: em_est11.*	55: em_est23.*
		12: em_est12.*	56: em_est24.*
		13: em_est13.*	57: em_est25.*
		14: em_est14.*	58: em_est26.*
		15: em_est15.*	59: gb_est33.*
		16: em_est16.*	60: gb_est34.*
		17: em_est17.*	61: gb_est35.*
		18: em_est18.*	62: gb_est36.*
		19: em_est19.*	63: gb_est37.*
		20: gb_est1.*	64: gb_est38.*
		21: gb_est2.*	65: em_est27.*
		22: gb_est3.*	66: em_est28.*
		23: gb_est4.*	67: em_est29.*
		24: gb_est5.*	68: em_est30.*
		25: gb_est6.*	69: gb_est39.*
		26: gb_est7.*	70: gb_est40.*
		27: gb_est8.*	71: gb_est41.*
		28: gb_est9.*	72: gb_est42.*
		29: gb_est10.*	73: gb_est43.*
		30: gb_est11.*	74: gb_est44.*
		31: gb_est12.*	75: em_est31.*
		32: gb_est13.*	76: em_est32.*
		33: gb_est14.*	77: em_est33.*
		34: gb_est15.*	78: em_est34.*
		35: gb_est16.*	79: gb_gss1.*
		36: gb_est17.*	80: gb_gss2.*
		37: gb_est18.*	81: gb_gss3.*
		38: gb_est19.*	82: gb_gss4.*
		39: gb_est20.*	83: em_gss1.*
		40: gb_est21.*	84: em_gss2.*
		41: gb_est22.*	85: em_gss3.*
		42: gb_est23.*	86: em_gss4.*
		43: gb_est24.*	87: gb_gss5.*
		44: gb_est25.*	88: gb_gss6.*
			89: gb_gss7.*
			90: gb_gss8.*
			91: gb_gss9.*
			92: em_gss5.*
			93: em_gss6.*
			94: em_gss7.*
			95: em_gss8.*
			96: em_gss9.*
			97: em_gss10.*
			98: em_gss11.*
			99: gb_gss10.*
			100: gb_gss11.*
			101: em_gss12.*
			102: gb_gss12.*
			103: gb_gss13.*
			104: gb_gss14.*
			105: gb_gss15.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match	Length	DB ID	Description	
c 1	47.8	8.0	925	79	CNS0091P	AL053013 Drosophil
2	43.8	7.4	1036	79	CNS010BS	AL098770 Drosophil

```
3 43.4 7.3 1065 79 CNS0108S
C 4 41.2 6.9 447 45 A858891 UI-R-A0-b
5 40.4 6.8 925 79 CNS0091P
C 6 40.2 6.8 550 49 A0057847 A0057847
7 38.8 6.5 162 79 FR0009521
8 38.8 6.5 623 33 AA413349 A000805 F.rubripe
9 38.4 6.5 472 72 AW162815 A000805 F.rubripe
10 38.4 6.5 472 72 AW163342 A000805 F.rubripe
C 11 38.4 6.5 571 44 A1260212 A000805 F.rubripe
12 38.4 6.5 1101 80 CNS0175Y A000805 F.rubripe
C 13 38.2 6.4 538 42 A114969 A000805 F.rubripe
14 38.2 6.4 814 42 A1161654 A000805 F.rubripe
C 15 38 6.4 436 38 A756319 A000805 F.rubripe
C 16 38 6.4 553 51 A1756282 A000805 F.rubripe
17 38 6.4 838 42 A1159665 A000805 F.rubripe
18 38 6.4 863 42 A1159667 A000805 F.rubripe
19 37.6 6.3 464 64 A0089826 A000805 F.rubripe
C 20 37.4 6.3 964 79 CNS003WG A000805 F.rubripe
21 37.2 6.3 304 62 A1920772 A000805 F.rubripe
C 22 37 6.2 529 42 A1159402 A000805 F.rubripe
23 37 6.2 602 34 A524163 A000805 F.rubripe
24 37 6.2 705 42 A110630 A000805 F.rubripe
25 37 6.2 705 51 AF063504 A000805 F.rubripe
C 26 37 6.2 947 79 CNS008C0 A000805 F.rubripe
C 27 37 6.2 1200 80 CNS01671 A000805 F.rubripe
28 36.8 6.2 596 79 FR0019514 A000805 F.rubripe
29 36.8 6.2 611 43 A0030115 A000805 F.rubripe
30 36.8 6.2 637 43 A0031310 A000805 F.rubripe
C 31 36.8 6.2 720 79 CNS00Y22 A000805 F.rubripe
32 36.6 6.2 439 47 A1492967 A000805 F.rubripe
33 36.6 6.2 520 51 A1727579 A000805 F.rubripe
34 36.6 6.2 542 51 A1730787 A000805 F.rubripe
35 36.6 6.2 543 51 A1730078 A000805 F.rubripe
36 36.6 6.2 546 51 A1726834 A000805 F.rubripe
37 36.6 6.2 556 51 A1729927 A000805 F.rubripe
38 36.6 6.2 563 70 A146257 A000805 F.rubripe
39 36.6 6.2 611 51 A1731740 A000805 F.rubripe
40 36.6 6.2 613 51 A1727110 A000805 F.rubripe
41 36.6 6.2 615 51 A1730728 A000805 F.rubripe
42 36.6 6.2 617 51 A1730726 A000805 F.rubripe
43 36.6 6.2 625 51 A1727201 A000805 F.rubripe
44 36.6 6.2 626 51 A1727196 A000805 F.rubripe
45 36.6 6.2 637 51 A1728230 A000805 F.rubripe
```

## ALIGNMENTS

```
RESULT 1
CNS0091P/c 925 bp DNA GSS
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL053013
VERSION AL053013.1 GI:4934461
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 925)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
```

melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Hammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

source  
1..925  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"  
/clone="BACR19D16"  
/note="end : TET3"

BASE COUNT 120 a 61 c 61 g 172 t 511 others  
ORIGIN

Query Match 8.0%; Score 47.8; DB 79; Length 925;  
Best Local Similarity 11.8%; Pred. No. 0.32;  
Matches 37; Conservative 151; Mismatches 125; Indels 0; Gaps 0;

```
QY 256 CCAGCCAGGACCGCGCTCTGCCGCGCGCGATCGGTCGACCGCGGAGC 315
Db 914 SBBSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
QY 316 AGATCATTTGCTAATCTAGTGAATCAGTGGGATTAATGTCACCAAGGAGCGTC 375
Db 854 NASSSCCGCGGABCMSCSSSSSCGASAGVKVRASGGAGKRGSGGASGSSSS 795
QY 376 AGACTGGAGCGCGCCACGACGAGGTCTACGGCGGTGACGCTCCACGACTACCGCCCG 435
Db 794 ACBSSSSSSCASGWSASSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 735
QY 436 CAGCGACAAAGTCCATACCGCGGACATGCGGACCTGTACGCTGGATGTTGGCAGCT 495
Db 734 SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 675
QY 496 AATAGTGGTGGCAGCGCGCTCTCTCCACGCGCGGCGGAGGAGGAGGAGGAG 555
Db 674 SMSAARSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 615
QY 556 AAGGCTCCGCGG 568
Db 614 SASSGMSSSSSSS 602
```

## RESULT 2

```
CNS010BS 1036 bp DNA GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BACN03H11 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL098770
VERSION AL098770.1 GI:5610381
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1036)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequençage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
```

library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.

```

FEATURES
source
pBelobaCl:
1. .1065
Location/Qualifiers
/organism="Drosophila"
/plasmid="pBelobaCl"
/ab_xref="taxon:722"
/clone_lib="DrosBAC"
/clone="BACN03J03"
/notes="end T7"
```

BASE COUNT	121 a	225 c	252 g	313 t	154 others
OBIGIN					

Query Match	7.3%	Score 43.4;	DB 79;	Length 1065;
Best Local Similarity	27.4%	Pred. No. 3.1;		
Matches	51;	Conservative	62;	Mismatches 73;
				Indels 0;
				Gaps 0;

**Qy** 141 AGTGGCGACGGCGAAACCGGGGTGTCCTCCTCGGGCACTACCGCGGCAGTAGCAGTCA 200  
:: | :: :: :: | : : ||| | : | | : | :: :: : : | :  
**Db** 755 RVAGSSSSSSSGSSGCTCTSTCKGGGSSSGCCGKGSSSKBGSSBKSGSGKGKT 814

[illegible][illegible]

QY 321 CATGC 326  
: |:

## RESULT 4

LOCUS	AA858891	447 bp	mRNA	EST
DEFINITION	UI-R-A0-bd-d-01-0-UI.s1		UI-R-A0 Rattus norvegicus CDNA clone	03-JUL-1999
DESCRIPTION	UI-R-A0-bd-d-01-0-UI.3		similar to gb AA422729 AA422729.v2	v29p009.s1

sequence.  
 AA858891  
 AA858891.1 GT:429084

KEYWORDS	EST.	
'SOURCE	Norway rat.	
ORGANISM	Rattus norvegicus	

REFERENCE

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 447)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

AUTHORS

**JOURNAL**  
**MEDLINE**  
**discovery**  
Genome Res. 6 (9), 791-806 (1996)  
9704477

Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa

431 Eastman Medical Research Building Iowa City, IA 52242  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: [msoares@blue.weeg.uiowa.edu](mailto:msoares@blue.weeg.uiowa.edu)

is likely internal to the message. cDNA Library Preparation: MM Oligo- $\alpha$  track not found, not 1 site shown in beginning of seq. Fatima Bernaldo, Ph.D. Clone distribution: clones will be available through Research Genetics. This clone is also available through

I.M.A.G.E. Consortium at LLNL ([infoimage.llnl.gov](http://infoimage.llnl.gov)). IMAGE ID=1776747 The following repetitive elements were found in this cdna sequence: 15-121, >(CAG)n#Simple\_repeat 61-179,

Fax: 319.335.3803  
 Email: msorensen@blue.weeg.uiowa.edu  
 Oligo-dT track not found, Not I site shown in beginning of sequence  
 is likely internal to the message. cDNA Library Preparation: M.  
 Fatima Bonaldo, Ph.D. Clone distribution: clones will be available  
 through Research Genetics This clone is also available through the  
 I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE  
 ID=1776747 The following repetitive elements were found in this  
 cDNA sequence: 15-121, >(CAG)n#Simple repeat 61-179,



Fax: 0298-38-7468  
Email: tsasaki@abr.affrc.go.jp  
PROJECT = 'RGP'

## FEATURES

Location/Qualifiers

1..550

/organism="Oryza sativa"

/strain="Nipponbare"

/db\_xref="taxon:4530"

/clone="S21840\_22"

/clone\_lib="Oryza sativa mature leaf Nipponbare"

/tissue\_type="mature leaf"

124 a 120 c 201 g 103 t 2 others

BASE COUNT

ORIGIN

## Query Match

Best Local Similarity 6.8%; Score 40.2; DB 49; Length 550;

Matches 84; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 148 ACGGGAACCGGGTGTGTCTCGGGCACTGACGGCGGAGTAGCAGTACATACCAAT 207

Db 309 ACGGTTTACCGGCTCAAGTCCCGACACGACGTACGCCGCTCTTCCCTTCGACGACGAG 250

Qy 208 AGCACTTTCGCGCTGCTGATCCCTGTCATCACCCCTGTTTCAGCGCCAGCAAGCA 267

Db 249 ACCCTTGTCTGCTGCTGATCCCGACACGTACGCCGCTCTTCCCTTCGACGACGAG 190

Qy 268 GCGGCGCTCTGCGCGGCGGATCGGCTGAGCCGC 304

Db 189 GCAGGCGCTGTTCTCCCTCGCCGACGAGCGCTCCTC 153

## RESULT 7

LOCUS

FR0009521 162 bp DNA GSS 18-SEP-1997

DEFINITION F.rubripes GSS sequence, clone 021G08aA1, genomic survey sequence.

ACCESSION AL000805

KEYWORDS AL000805.1 GI:2438657

SOURCE GSS: genome survey sequence.

Fugu rubripes.

Fugu rubripes

Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;

Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;

Tetraodontiformes; Tetraodontidae; Tetraodontidae; Fugu.

REFERENCE 1 (bases 1 to 162)

AUTHORS Williams, G. and Brenner, S.

TITLE Direct Submission

JOURNAL Submitted (09-SEP-1997) MRC Human Genome Mapping Project Resource

Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hmp.mrc.ac.uk

COMMENT V\_type: phagemid

PRIMER: KS

DESCR:

One pass dye-terminator sequencing of cosmid cloned genomic

sequence.

Location/Qualifiers

1..162

/organism="Fugu rubripes"

/db\_xref="taxon:31033"

/clone\_lib="cosmid 021G08"

/clone="021G08aA1"

8 a 54 c 43 g 50 t 7 others

BASE COUNT

ORIGIN

## Query Match

Best Local Similarity 6.5%; Score 38.8; DB 79; Length 162;

Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 170 CCTCGGCACTGACGGCGGAGTAGCAGTCATACCAATACGACTTCTGCGCTGCTGCAT 229

Db 3 CTNCGCGNCTGCTGNGCTACTNCTGCTCTACTACTGCTGCGNGCTGCTGCTA 62

Qy 230 CCCCTGCATCACCCCTGTTTTCAGCGCCAGCAAGGACGCCGCTCTCTGCGCGGCGC 289

Db 63 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 122

Qy 290 GATCGGCTG 298

Db 123 CTGCTGCTG 131

## RESULT 8

LOCUS

AA413349 623 bp mRNA EST 02-MAY-1997

DEFINITION AgEST00022 Anopheles gambiae adult, pSport cDNA Anopheles gambiae

cDNA clone cc29 5', mRNA sequence.

ACCESSION AA413349

VERSION AA413349.1 GI:2071903

KEYWORDS EST.

SOURCE African malaria mosquito.

ORGANISM

Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;

Culicoidae; Culicidae; Anopheles.

REFERENCE 1 (bases 1 to 623)

AUTHORS Cornet, A.J., Kumar, V., Mukabayire, O., Salazar Rafferty, C.,

Petrarca, V., Coluzzi, M. and Collins, F.H.

TITLE A comprehensive physical map of the malaria vector Anopheles

gambiae

JOURNAL

COMMENT Unpublished (1997)

On May 5, 1995 this sequence version replaced gi:797752.

Contact: Salazar Rafferty, C.; and Collins, F.H

Vector Genetics Section

Centers for Disease Control and Prevention

MS F-22, 4770 Buford Hwy, Chamblee, GA 30341

Tel: 770 488 7463

Fax: 770 488 7469

Email: czs7@cdc.gov

Mapping location: 23C in 2L chromosome

Seq primer: M13 Reverse

High quality sequence stop: 623.

FEATURES

source

1..623

/organism="Anopheles gambiae"

/strain="G3"

/db\_xref="taxon:7165"

/clone="cc29"

/clone\_lib="Anopheles gambiae adult pSport cDNA"

/sex="Male and Female"

/note="Site\_1: SalI; Site\_2: NotI; See: Salazar, C.E., et

al. Insect Molecular Biology (1994), 3:1-13."

BASE COUNT 154 a 196 c 221 g 52 t

ORIGIN

Query Match

Best Local Similarity 6.5%; Score 38.8; DB 33; Length 623;

Matches 94; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Qy 102 CCGCAGGGAAGAACACAGACGCCAGCCAAACATCGGTAGTGGGACCGGAAACCGG 161

Db 258 CCGGCGCGGGAACAGACAGCGCTCCGAGAGAGGACCGGCTGCGCGCGCTGC 317

Qy 162 GTGTGTCTCGGGCACTGACGGCGGAGTAGCAGTCATACCAATACGACTTCTGCCGC 221

Db 318 TGCTGAACCAAGCAAGCCTGCGGTGAGCGCAAGAGGATGCCAAGAGCTGCCCGGC 377

Qy 222 TGCTGCATCCCTGTCATCACCCCTGTTTCAGCGCCAGCAAGGACCGGCTCTCTGC 281

Db 378 CGCTGCTGCTGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 437

Qy 282 CGCGGC 287

Db 438 TGCTGC 443





```
RESULT 11
AI260212/c
LOCUS
DEFINITION
  AI260212 571 bp mRNA EST 17-NOV-1998
  LP03924.5prime LP Drosophila melanogaster larval-early pupal pot2
  Drosophila melanogaster cDNA clone LP03924 5prime similar to
  X61945: ng3 FBgn0010295 PID:9296042 SWISS-PROT:P40140, mRNA
  sequence.
ACCESSION
AI260212
VERSION
AI260212.1 GI:3867737
KEYWORDS
EST.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 571)
AUTHORS
Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
Brokstein,P., Lewis,S. and Rubin,G.M.
TITLE
BDGP/HMI Drosophila EST Project
JOURNAL
Unpublished (1997)
COMMENT
Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Sequence is the complete cDNA insert. hit genomic sequence AL024484
Plate: 39 row: B column: 12
High quality sequence stop: 513.
FEATURES
source
1..571
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="LP03924"
/sex="male and female"
/dev_stage="larvae-pupae"
/lab_host="DH5-alpha"
/notes="Organ: whole body; Vector: pot2; Site_1: EcoRI;
Site_2: XhoI; Sized fractionated cDNAs were directly
ligated into pot2. plasmid cDNA library."
BASE COUNT
161 a 169 c 114 g 126 t 1 others
ORIGIN
Query Match 6.5%; Score 38.4; DB 44; Length 571;
Best Local Similarity 56.2%; Pred. No. 34;
Matches 72; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
Qy 154 AACCGGGGTGTCTCTCGGCGACTGACGCGCGAGTAGCAGTACCAATACGACT 213
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 427 AGACAGCGCTCGGCACATCGGCTCTCTCTCGGCGGATGCGACGTTCACCACTCGGCT 368
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 214 TCTGCGGTGTGATCCCTGCATCACCCTGTTTCAGGCCAGCAAGCAGCCGCG 273
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 367 ATGTGTGTCACACTGCTGCTACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCT 308
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 274 CTCCTGCTC 281
| | | | |
Db 307 GCTGGTGC 300
| | | | |
RESULT 12
CNS017SY
LOCUS
DEFINITION
  CNS017SY 1101 bp DNA GSS 26-JUL-1999
  Drosophila melanogaster genome survey sequence sp6 end of BAC
  BACN37L08 of DrosBAC library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.
ACCESSION
AL108460
VERSION
AL108460.1 GI:5628764
```

```
KEYWORDS
fruit fly.
SOURCE
Drosophila melanogaster
ORGANISM
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1101)
AUTHORS
Genoscope.
TITLE
Direct Submission
JOURNAL
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT
- Web : www.genoscope.cns.fr)
- Determination of this BAC-end sequence was carried out as part of a
  collaboration with the European Drosophila Genome Project (EDGP) -
  http://www.edgp.ebi.ac.uk -- This Drosophila melanogaster BAC
  library (Dros BAC) was made by Alain Billaud at CEPH (Centre
  d'Etude du Polymorphisme Humain) with funding provided by a MRC
  project grant. The DNA was prepared from embryos by Alain Bucheton
  and Genevieve Payan. It has been constructed in the vector
  pBelOBAC11.
FEATURES
Location/Qualifiers
source
1..1101
/organism="Drosophila melanogaster"
/plasmid="pBelOBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN37L08"
/notes="end : SP6"
BASE COUNT
254 a 176 c 160 g 152 t 359 others
ORIGIN
Query Match 6.5%; Score 38.4; DB 80; Length 1101;
Best Local Similarity 15.1%; Pred. No. 40;
Matches 67; Conservative 182; Mismatches 195; Indels 0; Gaps 0;
Qy 137 CGGTAGTGGCAGCGCGAACCGGGTGTCTCTCGGCGACTGACGCGGAGTAGCA 196
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 638 CGGAGCGGRKGCACAKSAGMSCGGRSGSGGCGGKAKGVRGVRVCCAGGGASC 697
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 197 GTCATACCAATACGACTTCTCCGCTGCTGCATCCCTGCATCACCCTCTTTTCAGCGC 256
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 698 ACMAADCGCCAKMACCCSSSSSSSSSCASTSSASRGMVSSCACSGSGGASACGA 757
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 257 CAGCCAGGCGCGCGCTCTCTCGCGCGCGGATCGGCTGAGCCGACGTGGGAGCA 316
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 758 SGAGGSGRGGGCGCCASGVCGAACSSASCSSASMCGVSSGSCSASCSCCGVSSCA 817
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 317 AGATCATTTGCTAATCTAGTGAATCAGCTGGGATTAATGTACCCAAAGGAGCGTCA 376
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 818 VSASASVMKVASAVASCASAVASGMSAGAVSSSCRSVASSVSSAAASVSSVSSSSS 877
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 377 GCACGTGGAGCCCGCCACGACGAGGTCTACGGCGGTGACGTCCACGACTACCGCCCGC 436
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 878 SVVSAAVASASSSSASSASMAVAAAAVAVSSVAVSSVSSSSSSSSASVAVSVA 937
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 437 AGCGAACAGTCCATACGCGCACATGCCGACCTGTGACGCTGGATTGGTGGCAGTA 496
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 938 ASASVSSSSSSSSSTSSASVSSVSMGAVVSSSSSSSSSSSVSSVSSVAVAAASAAA 997
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 497 ATAGTGTGTCAGCGCGGCTCTCTCCACAGCGCGCGCAACCAACAGGAGAGAAA 556
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 998 AAAAAASSSASAVAVVSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSVSSV 1057
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 557 AGGCCTCCGCGCATGTGAACGA 580
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1058 AVASASASVSSVAAVAAVAAVAA 1081
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 13
AI114969/c
LOCUS
DEFINITION
  AI114969 538 bp mRNA EST 02-SEP-1998
  Sugano mouse embryo cdna clone Mus musculus
```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2000, 21:55:01 ; Search time 102.54 Seconds  
(without alignments)  
1449.328 Million cell updates/sec

Title: US-08-988-242-1\_COPY\_1232\_1825  
Sequence: 1 CAGGTACAGCGTAACGGCTT.....AAACGAGCTCGTGGCGATA 594

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	590.8	99.5	3402	1 T27310	Trypanosoma cruzi
2	39.2	6.6	2790	1 V18480	BOP1 CDNA, New Iso
3	38.2	6.4	2943	1 T16480	SAB virus gB glyco
4	36.8	6.2	3776	1 T42903	TRP-1 protein codi
5	36.6	6.2	203	1 V17226	SCA2 gene CAG repe
6	36.6	6.2	203	1 V30271	Glutamine rich reg
7	36.6	6.2	913	1 T07199	Cotton fibre-speci
8	36.6	6.2	913	1 T13034	Cotton fibre-speci
9	36.6	6.2	913	1 T30253	Cotton fibre-speci
10	36.6	6.2	913	1 T62610	Cotton fibre-speci
11	36.6	6.2	913	1 T70041	Cotton fibre-speci
12	36.6	6.2	1884	1 T13030	Cotton fibre-speci
13	36.6	6.2	1985	1 T30250	Cotton fibre-speci
14	36.6	6.2	1985	1 T70036	Cotton fibre-speci
15	36.4	6.1	2415	1 T86757	CDNA of the M3/6 g
16	36.4	6.1	2453	1 T86758	CDNA of the M3/6 g
17	36.2	6.1	633	1 Q03322	Genomic Eimeria te
18	36	6.1	795	1 V55830	FLGA insert stabl
19	36	6.1	799	1 V55831	Nucleotide sequenc
20	36	6.1	9600	1 V21683	Vector plasmid pCM
21	36	6.1	10596	1 Q51731	Plasmid pcisEBON f
22	36	6.1	10596	1 T40348	Plasmid pcisEBON f
23	36	6.1	10596	1 X15650	Nucleotide sequenc
24	35.8	6.0	3076	1 V43674	Receptor type tyro
25	35.6	6.0	1089	1 N92576	Sequence of the 1.
26	35.6	6.0	1089	1 T93593	Eimeria tenella sp
27	35.6	6.0	114955	1 X53491	Human adenosine Al
28	35.4	6.0	1463	1 V61483	Human secreted pro
29	35.2	5.9	417	1 Q25082	Antigen tc-11e gen
30	35.2	5.9	3376	1 Q75166	AF-9 CDNA, New acu
31	35	5.9	168	1 Q84833	Spinocerebellar at
32	35	5.9	171	1 Q84834	Spinocerebellar at
33	35	5.9	195	1 Q84831	Spinocerebellar at
34	35	5.9	234	1 Q84832	Spinocerebellar at

c 35 35 5.9 292 1 T75505 P. americanus skin  
c 36 35 5.9 627 1 V90494 EST clone DN740. N  
c 37 34.8 5.9 2001 1 T49282 DNA encoding B. la  
c 38 34.8 5.9 2001 1 T84788 Brevibacterium lac  
c 39 34.8 5.9 2001 1 V15788 B. lactofermentum  
c 40 34.8 5.9 2001 1 V40257 Brevibacterium lac  
c 41 34.8 5.9 6530 1 Q51557 Loridin gene, Con  
c 42 34.6 5.8 296 1 V89532 EST clone CP328. N  
c 43 34.6 5.8 1325 1 V80740 Human secreted pro  
c 44 34.6 5.8 2400 1 T42859 Choline oxidase ge  
c 45 34.6 5.8 2400 1 T75000 Choline oxidase ge

#### ALIGNMENTS

RESULT 1

T27310  
ID T27310 standard; cDNA; 3402 BP.  
AC T27310;  
DT 26-NOV-1996 (first entry)  
DE Trypanosoma cruzi epimastigotic Ptc100t antigen gene.  
KW Antigen; Trypanosoma cruzi; epimastigote; serum; Chagas disease; probe;  
KW Primer; PCR; polymerase chain reaction; amplification; antibody; ds.  
QS Trypanosoma cruzi.  
FH Key Location/Qualifiers  
FT CDS 266..3013  
FT /tag= a  
FT /product= PTC100t epimastigotic antigen

FR2723589-A1.  
16-FEB-1996.  
12-AUG-1994; 010132.  
12-AUG-1994; FR-010132.  
PA (INMR ) BIO MERIEUX.  
PI Jolivet M, Lesenechal M, Paranhos-Baccala G;  
DR WPI 96-190287/20.  
DR P-PSDB R91615.  
PT New nucleic acid encoding Trypanosoma cruzi epimastigotic antigen -  
PT useful for diagnosis, monitoring and therapy of Chagas disease  
PS Claim 1; Page 24-26; 5pp; French.  
CC This is the nucleotide sequence encoding a novel isolated antigenic  
CC protein from Trypanosoma cruzi epimastigotes, designated Ptc100t.  
CC The clone Tc50 was isolated from a T. cruzi genomic expression library in  
CC lambda gt11, using a mixture of sera from patients with Chagas disease  
CC Clone Tc50 contained an 594 bp insert corresp. to nucleotides 1232-1825  
CC of this sequence. The Tc50 sequence was subsequently used to probe a  
CC Southern blot of restriction enzyme digested T. cruzi DNA and also screen  
CC a lambda gt10 library to isolate a 1041 bp EcoRI fragment corresp. to  
CC nucleotides 1403-2443 of Ptc100t. Primers (T27311-5) were synthesised  
CC based on the sequences of the 594 and 1041 bp fragments and used to  
CC amplify the Ptc100t clone as 3 fragments from cDNA derived from mRNA  
CC purified from T. cruzi epimastigotes. The protein or antibodies raised  
CC against it can be used in the detection and monitoring of T. cruzi  
CC infection i.e. Chagas disease.  
SQ Sequence 3402 BP; 889 A; 818 C; 958 G; 737 T;

Query Match 99.5%; Score 590.8; DB 1; Length 3402;  
Best Local Similarity 99.7%; Pred. No. 1.2e-140;  
Matches 592; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGTACAGCGTAACGGCTTTTGGCTTCATCTGACAGCGAGGTAGCTCGCTCGGCT 60

DB 1232 CAGGTACAGCGTAACGGCTTTTGGCTTCATCTGACAGCGAGGTAGCTCGCTCGGCT 1291

QY 61 GACATGTCGATTCGATTCGATTCCTCCGTCCTCCGTCCTCCGTCCTCCGTCCTCCGTCCT 120

DB 1292 GACATGTCGATTCGATTCGATTCCTCCGTCCTCCGTCCTCCGTCCTCCGTCCTCCGTCCT 1351

QY 121 CCAGGCCAAAACATCGGTAGTGGCGAGCGGGAACACCGGGGTGTGTCTCTGGGCAT 180

DB 1352 CCAGGCCAAAACATCGGTAGTGGCGAGCGGGAACACCGGGGTGTGTCTCTGGGCAT 1411

QY	181	GAGCGCGAGTAGTACAGTATACCAATACGACTTCTGCCGCTGTGTCATCCCTGCATCA	240
Db	1412	GAGCGCGAGTAGTACAGTATACCAATACGACTTCTGCCGCTGTGTCATCCCTGCATCA	1471
QY	241	CCCCCTGTTTTCACGCGCCAGCCAGGACGCGCGCTCTCTGCCGCGGCGGCGATCGGCTGAG	300
Db	1472	CCCCCTGTTTTCACGCGCCAGCCAGGACGCGCGCTCTCTGCCGCGGCGGCGATCGGCTGAG	1531
QY	301	CCGACAGTGGGGACGAAGATCATTTCTAATCTAGTGAATACAGCTGGGGATTAATGTCAACC	360
Db	1532	CCGACAGTGGGGACGAAGATCATTTCTAATCTAGTGAATACAGCTGGGGATTAATGTCAACC	1591
QY	361	CAAGAGGAGCGTCTGTACGACTGGAGCGCGGCCACGACGAGGTTCTACGGCGGTGACGTCC	420
Db	1592	CAAGAGGAGCGTCTGTACGACTGGAGCGCGGCCACGACGAGGTTCTACGGCGGTGACGTCC	1651
QY	421	ACGACTACGCCCGCCGACGACCAAGTCCATACGGGGACAAATGCGCGACCTGTGACGGCT	480
Db	1652	ACGACTACGCCCGCCGACGACCAAGTCCATACGGGGACAAATGCGCGACCTGTGACGGCT	1711
QY	481	GGATTGGTGGCAGCTAATAGTGGTCCAGCGCGGCTCTCTCCACAGCGCGCGGCGAAA	540
Db	1712	GGATTGGTGGCAGCTAATAGTGGTCCAGCGCGGCTCTCTCCACAGCGCGCGGCGAAA	1771
QY	541	CCACAGGAGAGAAAAGCGCTCCGCGCATGTGAAACGAGCTCCGTGGCGATA	594
Db	1772	CCACAGGAGAGAAAAGCGCTCCGCGCATGTGAAACGAGCTCCGTGGCGATA	1825

RESULT 2

V18480

ID V18480 standard; cDNA to mRNA; 2790 BP.

AC V18480;

DT 18-AUG-1998 (first entry)

DE BOP1 cDNA.

KW Corticotroph; pituitary; BOP1; tumour suppressor; constitutive;

KW inhibible; Alzheimer's disease; nuclear transcription factor; apoptosis

KW cell cycle; neuronal disorder; ss.

OS Mus sp.

FH Key

FT Location/Qualifiers

FT 542..2545

FT CDS

FT /\*tag= a

FT /product= "BOP1 protein"

FN W09813489-A1.

PN 02-APR-1998.

PD 22-SEP-1997; E05198.

PF 23-SEP-1996; US-718661.

PR (CNRS ) CENT NAT RECH SCI.

PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

PI Journot L, Spengler D;

PI WPI: 98-230701/20.

DR P-PSDB; W48760.

PT New isolated tumour suppressor gene - useful for developing products

PT for use in diagnosis and treatment of tumour(s) or neuronal

PT disorder(s)

PS Claim 1; Pages 72-76; 118pp; English.

PS The present sequence represents the BOP1 cDNA isolated from the

CC mouse corticotroph pituitary tumour cell line Atr-20 cDNA library.

CC The protein encoded by the BOP1 cDNA displays a tumour suppressing

CC activity when it was constitutively and inducibly expressed in

CC tumour cells. The BOP1 cDNA and the protein it encodes are claimed

CC to be useful in the preparation of therapeutic compositions, useful

CC for treating, preventing or delaying the recurrence of a tumour or

CC neuronal disorders, e.g. genetic diseases or acquired degenerative

CC encephalopathies such as Alzheimer's disease. The BOP1 protein is

CC also claimed to be able to induce apoptosis resulting in inhibition

CC of tumour cell growth, to suppress tumour formation, to induce G1

CC arrest of the cell cycle and to act as nuclear transcription factor.

CC Sequence 2790 BP; 567 A; 783 C; 714 G; 626 T;

SQ

Query Match 6.6%; Score 39.2; DB 1; Length 2790;

Best Local Similarity 54.1%; Pred. No. 0.31;

	Matches	80;	Conservative	0;	Mismatches	68;	Indels	0;	Gaps
Qy	137	CGGTAGTGGCGAGCGGCAACACGGGGTGTGTCTCTCGGCACTGACCGCGGAGTAGCA	196						
Db	2114	CAGCAGCAGCCACCGGCATCGCGCGCCACACCGCTGCCACTGCCACAGCCACTGGCA	2173						
Qy	197	GTCATCAACATAGCACTTCTGCGCGTGTCTGCATCCCTGCATCAACCCCTGTTTCAGGCG	256						
Db	2174	GTGCCACAGCCACTACCAACAGCCACAGATGACGACACAGTTTCAGTTGCATCCAGGCC	2233						
Qy	257	CAGCCAAAGGAGCGCGGCTCTGCGCGC	284						
Db	2234	CAGATGCAGCTACACAGCTGCTGCGGC	2261						
RESULT	3								
TI6480/c									
ID	TI6480	standard; DNA; 2943 BP.							
AC	TI6480;								
DT	11-MAY-1996	(first entry)							
DE	SAB virus gB glycoprotein coding sequence.								
KW	Herpes simian monkey SAB virus gB glycoprotein; Immunoassay;								
KW	Diagnosis; herpes B virus; ss.								
OS	Herpes simian monkey SAB virus.								
FH	Key	Location/Qualifiers							
FT	cds	87..2744							
FT		/*tag- a							
FT		/product= gB_glycoprotein							
FN	US5487969-A.								
PD	30-JAN-1996.								
PF	01-APR-1993; 042747.								
PR	01-APR-1993; US-042747.								
PA	(SWBI-) SOUTHWEST FOUND BIOMEDICAL RES.								
FI	Black D, Eberle R, Hilliard J, Scinicariello F;								
DR	WPI; 96-105220/11.								
PT	P-PSDB; R92747.								
DT	Detection of herpes B virus by PCR amplification of sample DNA - to								
PT	detect a specific herpes simian monkey B virus DNA segment.								
PS	Disclosure; Column 23-30; 22pp; English.								
CC	This sequence encodes the herpes simian monkey SAB virus gB								
CC	glycoprotein. Such sequences have been used to develop synthetic								
CC	DNA primers which have homologous sequences of conserved regions								
CC	which flank a divergent region of the gB glycoprotein gene. In								
CC	addition, proteins such as the gB glycoprotein have potential use in								
CC	the development of serological immunoassays. One approach is to								
CC	synthesize peptides which, based on the properties of the predicted								
CC	protein sequence, are likely to be immunologically active. Such								
CC	peptides can be used as substrate antigens in immunoassays to detect								
CC	serum antibodies which recognize this specific peptide sequence.								
CC	Synthetic peptides may also be used to produce antibodies against								
CC	specific regions of the gB glycoprotein which are unique to one								
CC	virus. These can then be used to develop virus-specific								
CC	immunoassays for differentiation of SAB virus from other primate								
CC	alpha-herpes viruses and for identification of antibodies directed								
CC	against SAB virus in primate serum samples.								
SQ	Sequence 2943 BP; 462 A; 1103 C; 983 G; 395 T;								
Query Match		6.4%;	Score 38.2;	DB 1;	Length 2943;				
Best Local Similarity		47.7%;	Pred. No. 0,57;						
Matches 112;	Conservative	0;	Mismatches 123;	Indels	0;	Gaps			
Qy	209	CGACTTCTGCGGCTGCTGCATCCCTGCATCACCCCTGTTTCAGCGCCAGCCAGGACG	268						
Db	255	CGTCCTCTGTCGAGGACTCCTCTGCTCTCGGAAGCGCGGGCTGCGCGGGGTCGCGG	196						
Qy	269	CCGGCGCTCTGCGCGCGCGCATCGGCTGAGCCCGACGTGGGAGCAAGATCATCTCTA	328						
Db	195	TCGGGGCGCGCGCGCGCGGGGGGCTGTCGTCGGCGCGCGCGCATCATCGCGAGGCGGA	136						
Qy	329	ATCTAGTGAATCAGCTGGGGATTATGTTCACCAAGAGGAGGCTCTCAGCACTGGAGCGC	388						
Db	135	GGAGGACGGGAGGGGAAGAAAGGAGGGGGGGTCTCGGAGGCTCGATCTCCGGGGCGG	76						

















THIS PAGE BLANK (USPTO)